Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT®
Journal of Multivariate Analysis 92 (2005) 42–52

Journal of Multivariate Analysis

http://www.elsevier.com/locate/jmva

On the asymptotic properties of multivariate sample autocovariances

Georgi N. Boshnakov*

Mathematics Department, University of Manchester Institute of Science and Technology, P.O. Box 88, Manchester M60 1QD, UK

Received 14 January 2002

Abstract

We show that if a process can be obtained by filtering an autoregressive process, then the asymptotic distribution of sample autocovariances of the former is the same as the asymptotic distribution of linear combinations of sample autocovariances of the latter. This result is used to show that for small lags the sample autocovariances of the filtered process have the same asymptotic distribution as estimators utilizing more information (observations on the associated autoregression process and knowledge of the parameters of the filter). In particular, for a Gaussian ARMA process the first few sample autocovariances are jointly asymptotically efficient.

AMS 2000 subject classifications: 62M10; 62E20; 60F05

Keywords: Asymptotic efficiency; Multivariate ARMA; Serial covariances

1. Introduction

The sample autocovariances are widely used in time series analysis, they are easy to compute and it is of interest to know how efficient they are.

Under quite general conditions (see Section 5), any fixed vector of sample autocovariances is asymptotically normal and the elements of the asymptotic covariance matrix are given by the Bartlett's formulas plus, in the non-Gaussian case, terms involving fourth-order cumulants (for details see [1,5, Chapter 8]). In the

*Fax: +44-0-161-200-3669.
E-mail address: georgi.boshnakov@umist.ac.uk.

0047-259X/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2003.10.005

multivariate case the infinite sums in Bartlett's formulas can be replaced by the autocovariances corresponding to the tensor square of the spectral density matrix of the process (see [3]). For univariate processes the latter is simply the square of the spectral density (see [2,4,10]).

For Gaussian parametric models the Cramer–Rao bound for estimators of the autocovariances can be obtained by an appropriate transformation of the Cramer–Rao bound for the parameters. The latter can be found in [10] for the univariate case, and [7] for multivariate autoregressive moving average (ARMA) models. A direct comparison of the asymptotic covariance matrix of the sample autocovariances with the Cramer–Rao bound has been used by Porat [9] to show that the sample autocovariances for univariate ARMA(p,q) models are jointly asymptotically efficient up to lag $p-q$, provided that $p \geqslant q$, and are inefficient otherwise. Walker [11] derived this result by considering an approximation to the likelihood of the sample autocovariances. Porat [10] extended his earlier result (see [9]) to Gaussian processes other than ARMA by giving a necessary and sufficient condition for asymptotic efficiency of sample autocovariances in terms of the spectral density and its derivatives.

The method of Porat [10] was extended to multivariate processes by Kakizawa [6]. He did not consider joint efficiency for several lags although his method should be applicable to that problem too. Kakizawa was able to conclude that in the case of pure autoregressions of order p the autocovariances up to lag p are asymptotically efficient while for the pure moving average none of the autocovariances is asymptotically efficient. The mixed ARMA case remained open.

The efficiency of the sample autocovariances up to lag p for (univariate or multivariate) autoregressive models of order p, can be derived also indirectly from the efficiency of the least-squares and Yule–Walker estimators of the parameters of such models.

Comparisons with Cramer–Rao bound restrict the above methods to Gaussian processes. Our approach is to consider a class of models where the observed process is obtained by a finite linear transformation of a pure autoregressive process of order, say, p. In the univariate case this class of models is equivalent to the standard form of the ARMA model. In the multivariate case it provides one possible parameterization of the multivariate ARMA model.

We show that the sample autocovariances of the observed process have the same asymptotic distribution as (for brevity we say: are as good as) a linear combination of the sample autocovariances of the underlying autoregressive process (Theorem 1 and Corollary 1). Hence, when this linear combination involves only lags between 0 and p, the corresponding sample autocovariances of the observed process are as good as estimators that would be available if both, the underlying autoregression was observed, and the parameters of the linear combination were known. This result seems rather strong since the autocovariance and least-squares estimators for autoregressions are asymptotically equivalent. In the Gaussian case asymptotic efficiency of the sample autocovariances of the observed process follows from the fact that the first $p+1$ autocovariances of the underlying autoregressive process are asymptotically efficient (Theorems 2 and 3). The results are based on

Lemma 1 which establishes a weak equivalence property of the above-mentioned estimators.

2. The model

We consider weakly stationary d-variate processes with $d \geqslant 1$. Let Σ be a positive-definite matrix and $\mathscr{F}_t$ be the σ-field generated by ε_s, $s \leqslant t$, where the process $\{\varepsilon_t\}$ is such that

$$E(\varepsilon_t|\mathscr{F}_{t-1}) = 0, \quad E(\varepsilon_t\varepsilon_t'|\mathscr{F}_{t-1}) = \Sigma. \tag{2.1}$$

We assume that the observed process $\{y_t\}$ can be obtained by filtering a unobserved zero-mean d-variate autoregression process, i.e.,

$$x_t = \sum_{i=1}^{p} \phi_i x_{t-i} + \varepsilon_t, \tag{2.2}$$

$$y_t - \mu = x_t + \sum_{i=1}^{q} \theta_i x_{t-i} = \Theta' X_t, \tag{2.3}$$

where $\mu = Ey_t$; $q \geqslant 1$; $p \geqslant q$; ϕ_i, θ_j are $d \times d$ matrices for $i = 1, \ldots, p$, $j = 1, \ldots, q$; $X_t = \mathrm{Vec}(x_t, x_{t-1}, \ldots, x_{t-q})$, $\Theta' = (I, \theta_1, \ldots, \theta_q)$. Here Vec produces a vector by stacking the columns of its arguments one over the other. For symmetric matrices the analogous operator Vech omits the elements above the main diagonal.

Model (2.2)–(2.3) can be written more compactly as

$$\phi(B)x_t = \varepsilon_t, \tag{2.4}$$

$$y_t - \mu = \theta(B)x_t, \tag{2.5}$$

where $\phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$, $\theta(B) = 1 + \sum_{i=1}^{q} \theta_i B^i$. We assume that the zeroes of $\det(\phi(B))$ and $\det(\theta(B))$ are outside the unit circle. In the univariate case, $d = 1$, $\{y_t\}$ is an ARMA(p, q) process whose standard representation

$$y_t - \mu = \sum_{i=1}^{p} \phi_i(y_{t-i} - \mu) + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{2.6}$$

has autoregressive part $\phi(B)$ and moving average part $\theta(B)$. The inverse is also true, i.e., if a univariate ARMA process $\{y_t\}$ is defined by (2.6), then it can be represented as a filtered version of a pure autoregressive process by (2.2)–(2.3). In the multivariate case (2.2)–(2.3) provides an alternative parameterization of the ARMA model (see Section 7). We refer to $\phi(B)$ and $\theta(B)$ as the autoregressive and moving average polynomials, respectively.

Let $R_y(k) = E(y_t - \mu)(y_{t-k} - \mu)')$, $R_x(k) = E(x_t x_{t-k}')$, $k = 0, 1, \ldots$, be the autocovariance functions of $\{y_t\}$ and $\{x_t\}$, respectively. We do not lose anything by considering only nonnegative lags since $R_y(k) = R_y(-k)'$. Noting that X_t is a block vector with ith block $X_t^{(i)} = x_{t-i+1}$, $i = 1, \ldots, q+1$, and that

$E(y_t-\mu)(y_{t-k}-\mu)' = E(\Theta' X_t X_{t-k}'\Theta)$, we obtain the following relation between the autocovariance functions of $\{x_t\}$ and $\{y_t\}$:

$$R_y(k) = \Theta' Q_{k,q}\Theta, \quad k \geqslant 0, \tag{2.7}$$

where $Q_{k,q}$ is the block matrix

$$Q_{k,q} = \begin{pmatrix} R_x(k) & R_x(k+1) & \dots & R_x(k+q) \\ R_x(k-1) & R_x(k) & \dots & R_x(k+q-1) \\ \vdots & \vdots & \vdots & \vdots \\ R_x(k-q) & R_x(k-q+1) & \dots & R_x(k) \end{pmatrix}.$$

The sample autocovariances $\bar{R}_y(k)$ and $\bar{R}_x(k)$ can be defined by

$$\bar{R}_y(k) = \frac{1}{N}\sum_{t=k+1}^{N} (y_t - \bar{y}(N))(y_{t-k}' - \bar{y}(N)), \quad \bar{R}_x(k) = \frac{1}{N}\sum_{t=k+1}^{N} x_t x_{t-k}',$$

where $\bar{y}(N) = \frac{1}{N}\sum_{t=1}^{N} y_t$ and no centering is used for $\bar{R}_x()$ because $Ex_t = 0$. We denote by $\tilde{R}_y()$ estimators of $R_y()$ that would be available if the process $\{x_t\}$ was observed and Θ was known:

$$\tilde{R}_y(k) = \Theta' \bar{Q}_{k,q}\Theta, \tag{2.8}$$

where the matrix $\bar{Q}_{k,q}$ is defined as $Q_{k,q}$ with $\bar{R}_x()$ replacing $R_x()$.

3. The VeR operator

Define $\mathrm{VeR}(M) = \mathrm{Vech}(M)$ if matrix M is symmetric and $\mathrm{VeR}(M) = \mathrm{Vec}(M)$ otherwise. With more arguments, VeR is applied to each argument in turn and the result is combined in one vector: $\mathrm{VeR}(M_1, \dots, M_n) = \mathrm{Vec}(\mathrm{VeR}(M_1), \dots, \mathrm{VeR}(M_n))$.

We use VeR to define a vector of the non-redundant elements of $Q_{k,q}$:

$$R_x(k:k+q) = \mathrm{VeR}(R_x(\max(0,k-q)), \dots, R_x(k+q)),$$

where $k, q \geqslant 0$. The length, $v(k,q)$, of $R_x(k:k+q)$ is equal to $d(d+1)/2 + (k+q)d^2$ when $0 \leqslant k \leqslant q$, and to $(2q+1)d^2$ when $k > q$. We have also

$$R_x(0:k+q) = \mathrm{VeR}(R_x(0), \dots, R_x(k+q)),$$

$$R_x(s:s+q) = \mathrm{VeR}(R_x(\max(0,s-q)), \dots, R_x(s+q)).$$

So, for s between 0 and k the vector $R_x(s:s+q)$ is a subvector of $R_x(0:k+q)$ and therefore $R_x(s:s+q) = I(s,k,q)R_x(0:k+q)$, where $I(s,k,q)$ is the following matrix of zeroes and ones:

$$I(s,k,q) = \begin{cases} (I_{v(s,q)} \;\; 0_{v(s,q),v(0,k+q)-v(s,q)}) & \text{for } 0 \leqslant s \leqslant q, \\ (0_{v(s,q),m} \;\; I_{v(s,q)} \;\; 0_{v(s,q),v(0,k+q)-v(s,q)-m}) & \text{for } s > q. \end{cases}$$

Here $v(0,k+q)$ is the length of $R_x(0:k+q)$, $m = d(d+1/2) + (s-q-1)d^2$, and the size of $I(s,k,q)$ is $v(s,q) \times v(0,k+q)$.

The elements of the vector $\mathrm{VeR}(\Theta' Q_{k,q} \Theta)$ are linear combinations of the elements of $R_x(k:k+q)$. Its coefficients are quadratic functions of the elements of Θ and can be arranged in a matrix $\alpha_{k,q}(\Theta)$ such that

$$\mathrm{VeR}(\Theta' Q_{k,q} \Theta) = \alpha_{k,q}(\Theta) R_x(k:k+q), \tag{3.1}$$

see Appendix A for its explicit form. Vectors $R_y(k:k+q)$, $\bar{R}_x(k:k+q)$, $\bar{R}_y(k:k+q)$, $\tilde{R}_y(k:k+q)$ are defined similarly to $R_x(k:k+q)$ by replacing $R_x()$ with $R_y()$, $\bar{R}_x()$, $\bar{R}_y()$, and $\tilde{R}_y()$, respectively. Special cases are $R_x(k:k) = \mathrm{VeR}(R_x(k))$ and

$$\begin{aligned}&\sqrt{N}(\bar{R}_y(0:k) - R_y(0:k))\\ &\quad = \sqrt{N}\,\mathrm{VeR}(\bar{R}_y(0) - R_y(0), \bar{R}_y(1) - R_y(1), \ldots, \bar{R}_y(k) - R_y(k)).\end{aligned}$$

4. A lemma

In this section we show that $\bar{R}_y(k)$ and $\tilde{R}_y(k)$ are "close" even though $\tilde{R}_y(k)$ is based on more information (observations on the unobserved $\{x_t\}$ and knowledge of the moving average parameters Θ).

Lemma 1. *With the notation introduced so far the following relation holds*:

$$\bar{R}_y(k) = \tilde{R}_y(k) + q_N, \tag{4.1}$$

where q_N is such that $N^\gamma q_N \to 0$ in probability for any $\gamma \in (0,1)$.

Proof. It is sufficient to show that

$$\frac{1}{N} \sum_{t=k+1}^{N} (y_t - \mu)(y_{t-k} - \mu)' = \tilde{R}_y(k) + q_N. \tag{4.2}$$

Indeed, the variance of $\bar{y}(N)$ is $O(N^{-1})$ (because the spectral density of the process is continuous) and hence the variance of $N^{\gamma/2}\bar{y}(N)$ is $O(N^{\gamma-1})$. So, $N^{\gamma/2}(\bar{y}(N) - \mu)$ converges in mean square and in probability to zero. This is true also when the sum $\frac{1}{N}\sum_t y_t$ is from $k+1$ to N or from 1 to $N-k$. Hence, N^γ times the difference

$$\begin{aligned}&\bar{R}_y(k) - \frac{1}{N} \sum_{t=k+1}^{N} (y_t - \mu)(y_{t-k} - \mu)'\\ &\quad = -(\bar{y}(N) - \mu)\left(\frac{1}{N}\sum y_{t-k} - \mu\right)' - \left(\frac{1}{N}\sum y_t - \mu\right)(\bar{y}(N) - \mu)'\\ &\qquad + (\bar{y}(N) - \mu)(\bar{y}(N) - \mu)'\end{aligned}$$

converges in probability to zero (matrix), i.e., the replacement of the sample mean by the true mean μ introduces a term of the order required for q_N.

To show (4.2), we multiply $(y_t - \mu)$ by $(y_{t-k} - \mu)'$, use Eq. (2.3) and sum over t to get

$$\frac{1}{N}\sum_{t=k+1}^{N}(y_t-\mu)(y_{t-k}-\mu)' = \Theta'\left(\frac{1}{N}\sum_{t=k+1}^{N}X_tX_{t-k}'\right)\Theta. \tag{4.3}$$

The elements of the matrix in the parentheses are, up to end effects, sample autocovariances of the process x_t. Indeed, for $j>i$ the (i,j)th block of the term in parentheses in Eq. (4.3) is equal to

$$\begin{aligned}\frac{1}{N}\sum_{t=k+1}^{N}X_t^{(i)}(X_{t-k}^{(j)})' &= \frac{1}{N}\sum_{t=k+1}^{N}x_{t-i+1}x_{t-k-j+1}'\\ &= \frac{1}{N}\sum_{t=k+j-i+1}^{N}x_{t-i+1}x_{t-k-j+1}' + \frac{1}{N}\sum_{t=k+1}^{k+j-i}x_{t-i+1}x_{t-k-j+1}'\\ &= \bar{R}_x(k+j-i) + \frac{1}{N}\sum_{t=k+1}^{k+j-i}x_{t-i+1}x_{t-k-j+1}'.\end{aligned}$$

The factor $1/N$ ensures convergence in probability to zero of the second term on the last line above even after multiplication by N^γ, $\gamma\in(0,1)$, because the number of the summands in it is independent of N. The case $j<i$ is similar. Also, the diagonal $d\times d$ blocks (which correspond to $i=j$) of the term in parentheses in Eq. (4.3) are equal to $\bar{R}_x(k)$.

Since the (i,j)th block of the matrix $\bar{Q}_{k,q}$ is $\bar{R}_x(k+j-i)$ we get

$$\frac{1}{N}\sum_{t=k+1}^{N}X_tX_{t-k}' = \bar{Q}_{k,q} + O_p(N^{-1}).$$

Left-multiplying both sides of the last equation by Θ', right-multiplying by Θ and using the fact that the dimensions of all matrices involved are independent of N we get (4.2) and the required result. □

5. Equivalence of asymptotic distributions

The results below are valid whenever the sample autocovariances of the (unobserved) process $\{x_t\}$ are asymptotically normal. We do not need to impose any additional assumptions, this is emphasized by the formulations. In fact, the proof of Lemma 1 is valid under less restrictive conditions.

From Lemma 1 we get

$$N^\gamma(\bar{R}_y(k)-R_y(k)) - N^\gamma(\tilde{R}_y(k)-R_y(k)) = N^\gamma q_N = o_p(1). \tag{5.1}$$

Since this is an element-wise property, it holds also for the vectorized versions of the autocovariance matrices, namely

$$N^\gamma\,\mathrm{VeR}(\bar{R}_y(k)-R_y(k)) - N^\gamma\,\mathrm{VeR}(\tilde{R}_y(k)-R_y(k)) = o_p(1). \tag{5.2}$$

Hence, if $N^{\gamma}\,\mathrm{VeR}(\tilde{R}_y(k)-R_y(k))$ converges in distribution for some γ then so does $N^{\gamma}\,\mathrm{VeR}(\bar{R}_y(k)-R_y(k))$ and the two limit distributions coincide.

From Eqs. (2.7) and (2.8) we get

$$\begin{aligned} N^{\gamma}(\tilde{R}_y(k:k)-R_y(k:k)) &\equiv \mathrm{VeR}(N^{\gamma}(\tilde{R}_y(k)-R_y(k))) \\ &= \mathrm{VeR}(\Theta' N^{\gamma}(\bar{Q}_{k,q}-\bar{Q}_{k,q})\Theta) \\ &= \alpha_{k,q}(\Theta) N^{\gamma}(\bar{R}_x(k:k+q)-R_x(k:k+q)). \end{aligned} \tag{5.3}$$

Thus we have the following result.

Lemma 2. *Suppose that $N^{\gamma}(\bar{R}_x(k:k+q)-R_x(k:k+q))$ converges in distribution and let $k\geqslant 0$. Then the limiting distributions of*

$$N^{\gamma}\,\mathrm{VeR}(\bar{R}_y(k)-R_y(k)) \quad \textit{and} \quad \alpha_{k,q}(\Theta)N^{\gamma}(\bar{R}_x(k:k+q)-R_x(k:k+q))$$

exist and coincide.

Similar considerations show that the result of Lemma 2 holds jointly, i.e. that $N^{\gamma}\,\mathrm{VeR}(\bar{R}_y(0:k)-\bar{R}_y(0:k))$ and η have the same asymptotic distribution, where in obvious notation

$$\begin{aligned} \eta &= \mathrm{VeR}(\Theta'(\bar{Q}_{0,q}-Q_{0,q},\bar{Q}_{1,q}-Q_{1,q},\ldots,\bar{Q}_{k,q}-Q_{k,q})\Theta) \\ &= \begin{pmatrix} \alpha_{0,q}(\Theta)N^{\gamma}(\bar{R}_x(0:0+q)-R_x(0:0+q)) \\ \alpha_{1,q}(\Theta)N^{\gamma}(\bar{R}_x(1:1+q)-R_x(1:1+q)) \\ \cdots \\ \alpha_{k,q}(\Theta)N^{\gamma}(\bar{R}_x(k:k+q)-R_x(k:k+q)) \end{pmatrix} \\ &= \begin{pmatrix} \alpha_{0,q}(\Theta)I(0,k,q) \\ \alpha_{1,q}(\Theta)I(1,k,q) \\ \cdots \\ \alpha_{k,q}(\Theta)I(k,k,q) \end{pmatrix} N^{\gamma}(\bar{R}_x(0:k+q)-R_x(0:k+q)) \\ &= A(\Theta)N^{\gamma}(\bar{R}_x(0:k+q)-R_x(0:k+q)) \end{aligned} \tag{5.4}$$

($A(\Theta)$ is defined to be the first term on the penultimate line.) Notice that $A(\Theta)\bar{R}_x(0:k+q)$ is an estimator of $R_y(0:k)$ but only when $\{x_t\}$ is observed and Θ is known. The following theorem establishes that (asymptotically) the quality of $A(\Theta)\bar{R}_x(0:k+q)$ is the same as that of $\bar{R}_y(0:k)$ even though the former utilizes more information than the latter.

Theorem 1. *Suppose that $N^{\gamma}(\bar{R}_x(0:k+q)-R_x(0:k+q))$ converges in distribution and let $k\geqslant 0$. Then the limiting distributions of*

$$N^{\gamma}(\bar{R}_y(0:k)-R_y(0:k)) \quad \textit{and} \quad A(\Theta)N^{\gamma}(\bar{R}_x(0:k+q)-R_x(0:k+q))$$

exist and coincide.

Under quite general conditions the sample autocovariances of an autoregressive process are asymptotically normal. One set of "neat" conditions is obtained by assuming, in addition to (2.1), that $E(\varepsilon_t(i)\varepsilon_t(j)\varepsilon_t(k)|\mathscr{F}_{t-1})$ and $E(\varepsilon_t(i)\varepsilon_t(j)\varepsilon_t(k)\varepsilon_t(l)|\mathscr{F}_{t-1})$ exist and do not depend on t (see [5]). Here $\varepsilon_t(i)$ is the ith component of the vector ε_t. A particular case of this is the case of i.i.d. ε_t's having finite fourth moments.

The following corollary gives a formula for the covariance matrix in the asymptotic distribution of the sample autocovariances of the process $\{y_t\}$. This is valid whenever the autocovariances of the associated autoregression process are asymptotically normal. In particular, normality of the process is not required.

Corollary 1. *Let $k \geqslant 0$ and suppose that the sample autocovariances $\bar{R}_x(0), \bar{R}_x(1), \ldots, \bar{R}_x(k+q)$ of the pure autoregression model* (2.2) *are asymptotically normal. Let $N(0, \Gamma_x)$ be the asymptotic distribution of $\sqrt{N}(\bar{R}_x(0:k+q) - R_x(0:k+q))$. Then $\bar{R}_y(0), \bar{R}_y(1), \ldots, \bar{R}_y(k)$, are asymptotically normal and the asymptotic distribution of $\sqrt{N}(\bar{R}_y(0:k) - R_y(0:k))$ is $N(0, \Gamma_y)$, where $\Gamma_y = A(\Theta)\Gamma_x A(\Theta)'$.*

6. Efficiency of sample autocovariances

In this section we assume that the processes are Gaussian. In that case for the pure autoregressive process $\{x_t\}$ defined by Eq. (2.2), the first $p+1$ sample autocovariances $\bar{R}_x(k)$, $k = 0, 1, \ldots, p$, are jointly asymptotically efficient estimators of the corresponding theoretical autocovariances $R_x(k)$, $k = 0, 1, \ldots, p$, in the sense that the Cramer–Rao bound is achieved asymptotically. Moreover, for any lag $k > p$ the sample autocovariance $\bar{R}_x(k)$ is asymptotically strictly inefficient (see [9,10] for the univariate case and [6] for the multivariate case).

The joint asymptotic efficiency of $\bar{R}_x(k)$, $k = 0, 1, \ldots, p$, implies that (and is actually equivalent to) for any set of coefficients c_i, not all zero, the linear combination $\sum_{i=0}^{p} c_i \bar{R}_x(i)$ is an asymptotically efficient estimator of $\sum_{i=0}^{p} c_i \bar{R}_x(i)$. Also, for $0 \leqslant k \leqslant p-q$, $\tilde{R}_y(k)$ is a linear combination of $\bar{R}_x(k)$, $k = 0, 1, \ldots, p$ (see Eq. (2.8)). Hence, if the process $\{x_t\}$ is observed and Θ is known, then for $0 \leqslant k \leqslant p-q$, $\tilde{R}_y(k)$ is an asymptotically efficient estimate of $R_y(k) = \Theta' R_x(k:k+q)\Theta$.

$\tilde{R}_y(k)$ cannot be computed from observations on the process $\{y_t\}$ only but by Lemma 2 $\bar{R}_y(k)$ has the same asymptotic distribution as $\tilde{R}_y(k)$, the best estimator of $R_y(k)$ based on $\{x_t\}$ and Θ. But the best estimator based on $\{y_t\}$ only cannot be better than the best estimator based on $\{x_t\}$ and Θ, since the former can be obtained from the latter (see Eq. (2.3)). Hence, $\bar{R}_y(k)$ is asymptotically efficient for $0 \leqslant k \leqslant p-q$.

The joint efficiency of $\bar{R}_y(k)$, $k = 0, 1, \ldots, p-q$, follows similarly by noting that any linear combination of them has the same asymptotic distribution as a linear combination of $\bar{R}_x(0), \ldots, \bar{R}_x(p)$. We summarize the above in the following theorem.

Theorem 2. *If the process* $\{y_t\}$ *obeys the model defined by Eqs.* (2.2)–(2.3) *and* $\{\varepsilon_t\}$ *is Gaussian, then the sample autocovariances* $\bar{R}_y(k)$, $k = 0, 1, \ldots, p-q$, *are jointly asymptotically efficient.*

For lags k larger than $p-q$ the expression for $\bar{R}_y(k)$ in terms of $\bar{R}_x()$ (see (2.8) and (4.1)) contains lags of $\bar{R}_x(l)$ larger than p (such $\bar{R}_x(l)$ are not asymptotically efficient). Hence, better estimates can be obtained by replacing $\bar{R}_x(l)$ by a more efficient estimate. However, this is possible if $\{x_t\}$ is observed and θ is known. Otherwise, when only $\{y_t\}$ is observed, it is not immediately obvious whether $\bar{R}_y(k)$ can be bettered or whether it achieves its Cramer–Rao bound.

The maximal lag in Theorem 2 depends on p and q through $p-q$ only. Hence, efficiency still holds if p and q in (2.2)–(2.3) are allowed to vary.

Theorem 3. *If the process* $\{y_t\}$ *obeys the model defined by Eqs.* (2.2)–(2.3), *where the unknown* p *and* q *satisfy* $p-q \geqslant l \geqslant 0$, *and* $\{\varepsilon_t\}$ *is Gaussian, then the sample autocovariances* $\bar{R}_y(k)$, $k = 0, 1, \ldots, l$, *are jointly asymptotically efficient.*

In the univariate case the condition $p-q \geqslant l$ means that the difference between the number of poles and zeroes of the spectrum of the process is at least l.

7. Relation to ARMA models

Here we show that (2.4)–(2.5) provides an alternative way to parameterize the ARMA model, i.e., that every ARMA process obeys (2.4)–(2.5) and vice versa.

Let $\{y_t\}$ be an invertible multivariate ARMA process with representation

$$\alpha(B)(y_t - \mu) = \beta(B)\varepsilon_t, \tag{7.1}$$

where $\alpha(0) = \beta(0) = I$. Then $\beta^{-1}(B)\alpha(B)(y_t - \mu) = \varepsilon_t$ and there exist matrix polynomials $\phi(B)$ and $\theta(B)$ such that $\beta^{-1}(B)\alpha(B) = \phi(B)\theta^{-1}(B)$ and $\phi(0) = \theta(0) = I$. Hence, $\phi(B)\theta^{-1}(B)(y_t - \mu) = \varepsilon_t$. Now the substitution $x_t = \theta^{-1}(B)(y_t - \mu)$ gives (2.4)–(2.5).

Assume now that $\{y_t\}$ is governed by (2.4)–(2.5) and multiply both sides of Eq. (2.4) by $\theta(B)$ to get

$$\theta(B)\phi(B)x_t = \theta(B)\varepsilon_t. \tag{7.2}$$

This can be written also as $\theta(B)\phi(B)\theta^{-1}(B)\theta(B)x_t = \theta(B)\varepsilon_t$, which coupled with (2.5) gives $\theta(B)\phi(B)\theta^{-1}(B)(y_t - \mu) = \theta(B)\varepsilon_t$, where, in general, the operator on the left-hand side is not a polynomial in B. This can be remedied by left-multiplying both sides of the equation by an appropriate matrix polynomial, $\psi(B)$, to get $\psi(B)\theta(B)\phi(B)\theta^{-1}(B)(y_t - \mu) = \psi(B)\theta(B)\varepsilon_t$, so that $\{y_t\}$ is indeed an ARMA process.

8. Conclusion

By parameterizing the ARMA model as filtered autoregression we have shown that the sample autocovariances $\bar{R}_y(0), \bar{R}_y(1), \ldots, \bar{R}_y(p-q)$ of the observed process $\{y_t\}$ have the same asymptotic distribution as estimators based on the sample autocovariances of the associated unobserved autoregression $\{x_t\}$. Loosely speaking, we would not get better estimators of $R_y(0), R_y(1), \ldots, R_y(p-q)$ even if we were able to observe $\{x_t\}$ and knew the parameters of the filter relating $\{x_t\}$ and $\{y_t\}$.

Using the autoregression process as a benchmark is natural since least squares and, more generally, conditional least-squares estimators for autoregressions are routinely used and are asymptotically equivalent to autocovariance estimators under quite general conditions. In the particular case when the innovations are independent Gaussian, the sample autocovariances $\bar{R}_y(0), \bar{R}_y(1), \ldots, \bar{R}_y(p-q)$ are jointly asymptotically efficient.

Acknowledgments

I would like to thank the referees for their constructive comments which helped to improve this paper.

Appendix A. An expression for $\alpha_{k,q}(\Theta)$

We expand the right-hand side of Eq. (2.7) using the block forms of the participating matrices (see also the line after Eq. (2.3)) and arrange for the argument of $R_x()$ to be a summation index:

$$\begin{aligned}\Theta' Q_{k,q}\Theta &= \sum_{l=0}^{q}\sum_{j=0}^{q}\theta_j R_x(k+l-j)\theta_l' \\ &= \sum_{m=k}^{k+q}\sum_{j=0}^{k+q-m}\theta_j R_x(m)\theta_{m-k+j}' + \sum_{m=k-q}^{k-1}\sum_{j=k-m}^{q}\theta_j R_x(m)\theta_{m-k+j}' \\ &= \sum_{m=k}^{k+q}\sum_{j=0}^{k+q-m} + \sum_{m=\max(k-q,0)}^{k-1}\sum_{j=k-m}^{q} + \sum_{m=k-q}^{-1}\sum_{j=k-m}^{q},\end{aligned}$$

where, by convention, a sum is zero if the lower limit is greater than the upper limit. The third term above involves negative lags of the autocovariance function and is present only if $k<q$. We have (see [8, p. 30]) $\operatorname{Vec}(\theta_j R_x(m)\theta_{m-k+j}') = (\theta_{m-k+j}\otimes\theta_j)\operatorname{Vec}(R_x(m))$. Since $R_x(m) = R_x(-m)'$, it follows also that $\operatorname{Vec}(R_x(m)) = K_{dd}\operatorname{Vec}(R_x(-m))$, where K_{dd} is a commutation matrix (see [8, p. 47]). This shows that, for $k-q\leqslant m\leqslant -1$,

$\mathrm{Vec}(\theta_j R_x(m)\theta'_{m-k+j}) = (\theta_{m-k+j}\otimes\theta_j)K_{dd}\,\mathrm{Vec}(R_x(-m))$. Let

$$A_m = \begin{cases} \sum_{j=0}^{k+q-m}(\theta_{m-k+j}\otimes\theta_j) & \text{for } m = k,\ldots,k+q,\\ 0 & \text{otherwise,}\end{cases}$$

$$B_m = \begin{cases} \sum_{j=k-m}^{q}(\theta_{m-k+j}\otimes\theta_j) & \text{for } m = \max(k-q,0),\ldots,k-1,\\ 0 & \text{otherwise,}\end{cases}$$

$$C_m = \begin{cases} \sum_{j=k+m}^{q}(\theta_{-m-k+j}\otimes\theta_j)K_{dd} & \text{for } m = 1,\ldots,q-k \text{ and } q>k,\\ 0 & \text{otherwise.}\end{cases}$$

Let also $G_m = A_m + B_m + C_m$, $V_m = \mathrm{Vec}(R_x(m))$. Then,

$$\begin{aligned}\mathrm{Vec}(\Theta' Q_{k,q}\Theta) &= \sum_{m=k}^{k+q} A_m V_m + \sum_{m=\max(k-q,0)}^{k-1} B_m V_m + \sum_{m=1}^{q-k} C_m V_m\\ &= \sum_{m=\max(k-q,0)}^{k+q} (A_m + B_m + C_m)\,\mathrm{Vec}(R_x(m))\\ &= (G_{\max(k-q,0)}\ \ldots\ G_{k+q})R_x(k:k+q). \qquad \text{(A.1)}\end{aligned}$$

Let D_d be the duplication matrix (see [8, p. 49]). For $k=0$, Eq. (A.1) gives $\mathrm{Vech}(\Theta' Q_{0,q}\Theta) = D_d^+(G_0\ \ldots\ G_{k+q})R_x(0:0+q)$. From this and (A.1) we get Eq. (3.1) by putting

$$\alpha_{k,q}(\Theta) = \begin{cases} D_d^+(G_0\ \ldots\ G_{k+q}) & \text{for } k=0,\\ (G_{\max(k-q,0)}\ \ldots\ G_{k+q}) & \text{for } k>0.\end{cases}$$

References

[1] T.W. Anderson, The Statistical Analysis of Time Series, Wiley, New York, 1971.
[2] G.N. Boshnakov, On the asymptotic distribution of the sample autocovariance and autocorrelation functions, C. R. Acad. Bulgare Sci. 42 (5) (1989) 21–23.
[3] G.N. Boshnakov, The asymptotic covariance matrix of the multivariate serial correlations, Stochastic Process. Appl. 65 (2) (1996) 251–258.
[4] G.N. Boshnakov, Bartlett's formulae—closed forms and recurrent equations, Ann. Inst. Statist. Math. 48 (1) (1996) 49–59.
[5] E.J. Hannan, The asymptotic distribution of the serial covariances, Ann. Statist. 4 (1976) 396–399.
[6] Y. Kakizawa, Note on the asymptotic efficiency of sample autocovariances in Gaussian vector stationary processes, J. Time Ser. Anal. 20 (5) (1999) 551–558.
[7] A. Klein, A generalization of the Whittle's formula for the information matrix of vector-mixed time series, Linear Algebra Appl. 321 (2000) 197–208.
[8] J.R. Magnus, H. Neudecker, Matrix differential calculus with applications in statistics and econometrics, Revised Edition, Wiley Series in Probability and Statistics, Wiley, Chichester, 1999.
[9] B. Porat, Some asymptotic properties of the sample covariances of Gaussian autoregressive moving-average processes, J. Time Ser. Anal. 8 (1987) 205–220.
[10] B. Porat, Digital Processing of Random Signals: Theory and Methods, Prentice-Hall, Englewood Cliffs, NJ, 1994.
[11] A.M. Walker, On results of Porat concerning asymptotic efficiency of sample covariances of Gaussian ARMA processes, J. Time Ser. Anal. 16 (2) (1995) 237–248.